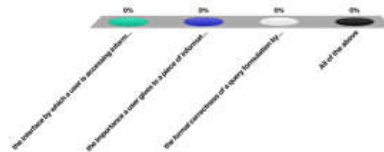


## A retrieval model attempts to model ...

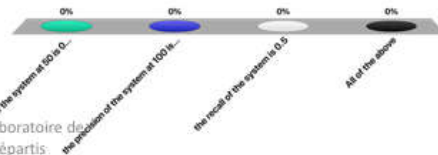
1. the interface by which a user is accessing information
2. the importance a user gives to a piece of information
3. the formal correctness of a query formulation by user
4. All of the above



© 2018, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis

Q: If the top 100 documents contain 50 relevant documents ...

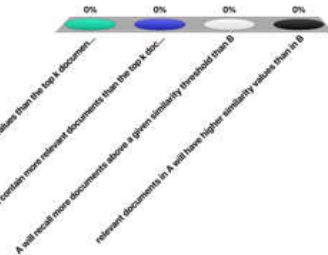
1. the precision of the system at 50 is 0.25
2. the precision of the system at 100 is 0.5
3. the recall of the system is 0.5
4. All of the above



© 2018, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis

## If retrieval system A has a higher precision than system B ...

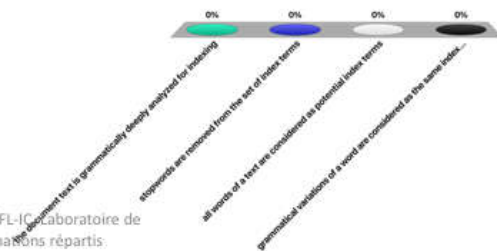
1. the top k documents of A will have higher similarity values than the top k documents of B
2. the top k documents of A will contain more relevant documents than the top k documents of B
3. A will recall more documents above a given similarity threshold than B
4. relevant documents in A will have higher similarity values than in B



©2018, Karl Aberer, EPFL-IG, Laboratoire de systèmes d'informations répartis

## Full-text retrieval refers to the fact that ...

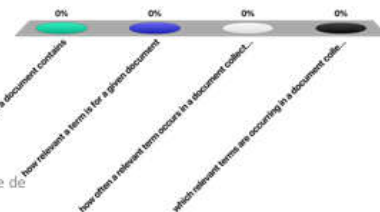
1. the document text is grammatically deeply analyzed for indexing
2. stopwords are removed from the set of index terms
3. all words of a text are considered as potential index terms
4. grammatical variations of a word are considered as the same index terms



© 2018, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis

## The entries of a term-document matrix indicate ...

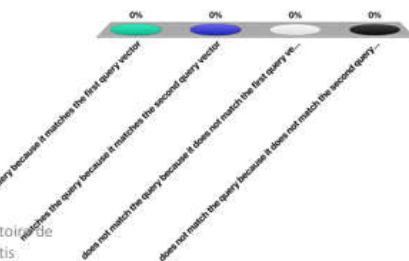
1. how many relevant terms a document contains
2. how relevant a term is for a given document
3. how often a relevant term occurs in a document collection
4. which relevant terms are occurring in a document collection



© 2018, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis

Let the query be represented by  $\{(1, 0, -1), (0, -1, 1)\}$  and the document by  $(1, 0, 1)$ . The document ...

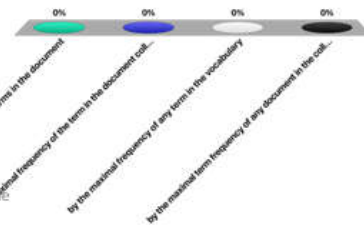
1. matches the query because it matches the first query vector
2. matches the query because it matches the second query vector
3. does not match the query because it does not match the first query vector
4. does not match the query because it does not match the second query vector



© 2018, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis

## The term frequency of a term is normalized ...

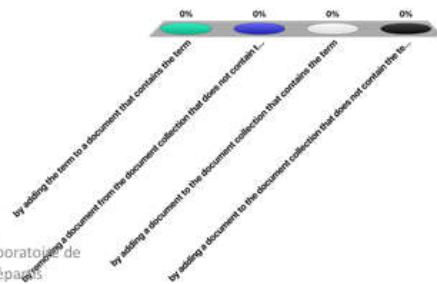
1. by the maximal frequency of all terms in the document
2. by the maximal frequency of the term in the document collection
3. by the maximal frequency of any term in the vocabulary
4. by the maximal term frequency of any document in the collection



© 2018, Karl Aberer, EPFL-ICAR Laboratoire  
systèmes d'informations réparties

## The inverse document frequency of a term can increase ...

1. by adding the term to a document that contains the term
2. by removing a document from the document collection that does not contain the term
3. by adding a document to the document collection that contains the term
4. by adding a document to the document collection that does not contain the term



© 2018, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis



Consider the document:

“Information retrieval is the task of finding the documents satisfying the information needs of the user”

Using MLE to estimate the unigram probability model, what is  $P(\text{the} | M_d)$  and  $P(\text{information} | M_d)$ ?

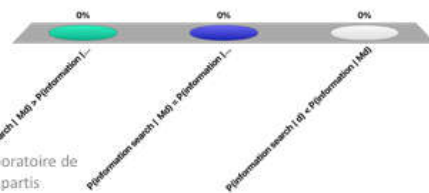
1.  $1/16$  and  $1/16$
2.  $1/12$  and  $1/12$
3.  $1/4$  and  $1/8$
4.  $1/3$  and  $1/6$



Consider the following document

$d = \text{"information retrieval and search"}$

1.  $P(\text{information search} \mid M_d) > P(\text{information} \mid M_d)$
2.  $P(\text{information search} \mid M_d) = P(\text{information} \mid M_d)$
3.  $P(\text{information search} \mid d) < P(\text{information} \mid M_d)$



© 2018, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis